



# Probabilistic model checking of biological systems with uncertain kinetic rates

Roberto Barbuti<sup>a</sup>, Francesca Levi<sup>a</sup>, Paolo Milazzo<sup>a</sup>, Guido Scatena<sup>b,\*</sup>

<sup>a</sup> Dip. di Informatica, Univ. di Pisa, Largo B. Pontecorvo 3, 56127 - Pisa, Italy

<sup>b</sup> IMT Lucca Inst. for Advanced Studies, Piazza San Ponziano 6, 55100 - Lucca, Italy

## ARTICLE INFO

### Article history:

Received 16 May 2011

Received in revised form 5 October 2011

Accepted 22 October 2011

Communicated by J. Hillston

### Keywords:

Probabilistic model checking

Systems biology

Uncertain kinetic rates

Abstract interpretation

Interval Markov chains

## ABSTRACT

In this paper, we present a formalization of biological systems based on multiset rewriting and we investigate the use of abstract interpretation on its semantics. We consider a probabilistic semantics, which is well suited to represent the non-deterministic evolution of real biological systems. Abstract interpretation allows us to deal with systems in which the kinetic rates of the evolution rules are not precisely known. On the (abstract) systems, we perform probabilistic model checking obtaining lower and upper bounds for the probabilities of reaching states satisfying the given properties. We apply abstract probabilistic model checking to verify reachability properties in a model of tumor growth.

© 2011 Elsevier B.V. All rights reserved.

## 1. Introduction

Modeling biological systems requires representation of the events (reactions) which guide the evolution of the systems together with their rates. Rates are often not precisely known, given the difficulty of measuring them for each single reaction. Thus, in many cases, it is necessary to construct models with some approximation which should not influence the overall behavior of the system we are interested to analyze. In these cases, we can predict the evolution of the system, although in a non-precise way.

In this paper, that is an extended and revised version of [5], we present a formalization of biological systems based on *Multiset Rewriting* (MSR) [8], and we investigate the use of abstract interpretation [12] on its semantics with the aim of validating probabilistic temporal properties.

We choose MSR because it is simple and expressive enough to describe many systems of interest. Moreover, as many formalisms used in the context of biological systems modeling are based on MSR, techniques developed for it may be further adapted to more complex languages [39,4,11].

We consider a probabilistic semantics of MSR which is well suited to represent the non-deterministic evolution of real biological systems. Such a semantics is given in terms of *Discrete Time Markov Chain* (DTMC) which can be used to perform probabilistic model checking. We define an effective method to compute an approximation of the probabilistic semantics of MSR systems for which the exact kinetic rates are not precisely known, but they are supposed to lie in some intervals. We use an *Interval Markov Chain* (IMC) [27,33] to abstract the set of DTMCs describing the probabilistic semantics of a set of MSR systems with uncertain kinetic rates. IMC is a model which combines non deterministic and probabilistic steps, using intervals of probabilities. Probabilistic model checking on IMC, which can be realized following the approach of [19], reports

\* Corresponding author. Tel.: +39 3391682585.

E-mail addresses: [barbuti@di.unipi.it](mailto:barbuti@di.unipi.it) (R. Barbuti), [levi@di.unipi.it](mailto:levi@di.unipi.it) (F. Levi), [milazzo@di.unipi.it](mailto:milazzo@di.unipi.it) (P. Milazzo), [g.scatena@imtlucca.it](mailto:g.scatena@imtlucca.it) (G. Scatena).

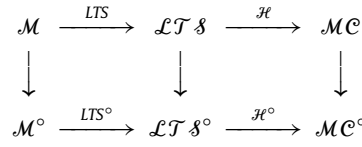


Fig. 1. Schematics of the defined theory.

lower and upper bounds for probabilistic temporal properties. In particular, we are interested in probability of reachability properties, that is in the probability to reach states satisfying given properties.

We start by recalling MSR. MSR is used as the formalism for constructing *concrete systems*, namely systems with exact kinetic rates. We give a *Labeled Transition System* (LTS) semantics to MSR and show how to derive, in standard way, a probabilistic semantics from it, in terms of a DTMC.

In order to deal with uncertainty we define *abstract systems*, in which the kinetic rates are given as intervals, we introduce an abstract LTS semantics and a systematic method to derive an IMC from abstract LTS.

In order to prove the soundness (and precision) of the proposed approach w.r.t. probabilistic reachability we use abstract interpretation concepts [12]. To this aim, we relate the abstract semantics, both LTS and probabilistic, to their concrete versions by means of abstraction functions, as schematically shown in Fig. 1. Here  $\mathcal{M}$ ,  $\mathcal{LTS}$  and  $\mathcal{MC}$  denote the domains of concrete systems, LTSs and DTMCs, respectively, and  $^o$  denotes their abstract versions. Moreover, the functions  $LTS$  and  $\mathcal{H}$  define the LTS semantics and the translation into DTMC, respectively, and  $LTS^o$  and  $\mathcal{H}^o$  define their abstract versions. Abstraction functions, represented by vertical arrows, allow us to formally specify the set of concrete elements represented by an abstract element in the associated domain, in standard abstract interpretation style.

We prove the soundness (and precision) of the approach by showing that: (i) an abstract LTS coincides with the *most precise* abstraction of the set of LTS it represents; (ii) analogously, an IMC coincides with the *most precise* abstraction of the set of DTMC it represents. This guarantees that lower and upper bounds of probabilistic reachability, computed on the IMC of an abstract system, are exactly the *most precise values* which are correct. Indeed, they correspond to the minimum and the maximum of the exact probabilities for probabilistic reachability, calculated over the DTMC for each concrete system represented by the abstract one.

To validate the usefulness of our approach in the context of biological systems modeling, we apply probabilistic model checking to verify reachability properties in an abstract system of tumor growth [43]. We conclude with a discussion about related works and we present some possible future research directions.

## 2. Probabilistic model checking of biological systems

To model biological systems we adopt MSR [8] where rewriting rules are enriched with non negative real kinetic constants. In this model, multisets are states of computation and transitions between states are obtained by applying rewriting rules with a probability proportional to their kinetic constants.

Let  $\Sigma$  be a finite set of *species names*. A *multiset* is a function  $s : \Sigma \mapsto \mathbb{N}$  and  $\mathcal{S}(\Sigma)$  is the *universe of multisets over  $\Sigma$* . Multiset sum  $\oplus$  and difference  $\ominus$  are defined as follows: for  $s', s'' \in \mathcal{S}(\Sigma)$ , we have  $s' \oplus s''(x) = s'(x) + s''(x)$  and  $s' \ominus s''(x) = \max(s'(x) - s''(x), 0)$ . In what follows we shall often assume  $\Sigma$  to be given.

A multiset represents a configuration of a biological system and possible events are modeled by rewriting rules. A *rewriting rule* is a pair  $R = (l, r)$ , where  $l \in \mathcal{S}(\Sigma)$  and  $r \in \mathcal{S}(\Sigma)$  are multisets, called *reactants* and *products*, respectively. Each rule is associated to a *kinetic constant* that is, roughly, an indication of the likelihood of the represented event.

**Definition 1** (Concrete System). A concrete system  $M$  is a triple  $(\mathcal{R}, \mathcal{K}, s_0)$ :

- $\mathcal{R} = \{R_1, \dots, R_m\}$  is a vector of *rewriting rules*, where  $R_i \in \mathcal{S}(\Sigma) \times \mathcal{S}(\Sigma)$  for  $i \in \{1, \dots, m\}$ ;
- $\mathcal{K} = \{k_1, \dots, k_m\}$  is a vector of *kinetic constants*, where  $k_i \in \mathbb{R}_{\geq 0}$  for  $i \in \{1, \dots, m\}$ ;
- $s_0 \in \mathcal{S}(\Sigma)$  is the *starting state*.

In what follows we refer to generic tuples components by name. For instance, given a system  $M = (\mathcal{R}, \mathcal{K}, s_0)$ , we use  $\mathcal{R}(M)$ ,  $\mathcal{K}(M)$ ,  $s_0(M)$  to denote  $\mathcal{R}$ ,  $\mathcal{K}$ ,  $s_0$  respectively. When  $M$  is clear from the context, for  $i \in \{1, \dots, m\}$ , we use  $l_i$  and  $r_i$  to denote the reactants and the products multisets of the  $i$ -th rule  $\mathcal{R}[i]$ . Similarly, we use  $k_i$  for the  $i$ -th kinetic constant  $\mathcal{K}[i]$ .

The universe of concrete systems is denoted by  $\mathcal{M}$ . We also say that two concrete systems  $M_i, i \in \{1, 2\}$ , are *isomorphic* ( $M_1 \sim M_2$ ) if and only if  $\mathcal{R}(M_1) = \mathcal{R}(M_2) \wedge s_0(M_1) = s_0(M_2)$ . Intuitively,  $M_1 \sim M_2$  iff  $M_1$  and  $M_2$  share the initial state and the set of rewriting rules.

### 2.1. Labeled transition system semantics

To describe the semantics of a concrete systems we adopt a *Labeled Transition System* (LTS) semantics. Namely, we adopt a *transition relation* of the form  $s' \xrightarrow{\eta, \beta} s''$ , where  $\eta$  is the number of the applied rule and  $\beta \in \mathbb{R}_{\geq 0}$  is the *transition rate*.

The application of a rule  $R_\eta$  to a state  $s'$  is modeled by the inference rule

$$\boxed{\frac{\mathcal{R}[\eta] = (l, r) \quad \mathcal{K}[\eta] = k \quad l \subseteq s' \quad \beta = \text{rate}(l, s', k) \quad s'' = ((s' \ominus l) \oplus r)}{s' \xrightarrow{\eta, \beta} s''}} \quad (1)$$

where  $\text{rate}(l, s', k) = \text{kin}(l, s') \times k_\eta$  and  $\text{kin}(l, s') = \prod_{x \in \Sigma} \binom{s'(x)}{l(x)}$ .

To compute  $\text{kin}(l_\eta, s')$  we take into account the number of possible distinct applications of the rule  $R_\eta$  to state  $s'$ . Actually, this requires to compute the number of distinct combinations of the reactants  $l_\eta$  in the multiset  $s'$ . Then,  $\text{rate}(l_\eta, s', k_\eta)$  is obtained by multiplying the value of  $\text{kin}(l_\eta, s')$  by the kinetic constant  $k_\eta$  associated with rule  $R_\eta$ . Function  $\text{rate}$  is inspired by the mass action kinetics of chemical reactions.

In the following, we use  $\mathcal{LTS}$  to denote the universe of LTSs and we define the function  $LTS : \mathcal{M} \mapsto \mathcal{LTS}$ , such that  $LTS(M)$ , with  $M = (\mathcal{R}, \mathcal{K}, s_0)$ , is the LTS  $(S, s_0, \rightarrow)$ , obtained by transitive closure of (1) starting from  $s_0$ .

When the transition relation  $\rightarrow$  is clear from the context, we use  $Next(s)$  for the set of transitions exiting from the state  $s$ . In addition, we use  $TS(s', s'') = \{s' \xrightarrow{\eta, \beta} s'' \text{ for some } \eta, \beta\}$  to denote the set of transitions from  $s'$  to  $s''$ . Given a transition  $t = s' \xrightarrow{\eta, \beta} s''$ , we also use  $\text{rate}(t) = \beta$ . Note that,  $\forall R_\eta \in \mathcal{R}, s \in S$ , there is at most one transition  $s \xrightarrow{\eta, \beta} s' \in Next(s)$  corresponding to  $R_\eta$ .

## 2.2. Derivation of probabilistic semantics

We define the probabilistic semantics of a concrete system by means of a translation from its LTS into a DTMC.

Given a countable set  $S$  we denote by  $Distr(S) = \{\rho \mid \rho : S \mapsto [0, 1] \wedge \sum_{s \in S} \rho(s) = 1\}$  the set of *probability distributions* and with  $PDistr(S) = \{\rho \mid \rho : S \mapsto [0, 1]\}$  the set of *probability pseudo-distributions*.

**Definition 2** (*Discrete Time Markov Chain*). A DTMC is a triple  $(S, s_0, P)$ , where:

- $S$  is the set of states,  $s_0 \in S$  is the starting state;
- $P : S \mapsto Distr(S)$  is the transition probability function.

In a DTMC  $P(s, s')$  reports the probability of moving from state  $s$  to state  $s'$ . In the following, we restrict our attention to *finitely branching* DTMCs, meaning that for each  $s$  in the state space, the set  $\{s' \mid P(s, s') > 0\}$  is finite. Since our systems have  $m$ -sized vector of rules, for each state, we have at most  $m$  outgoing transitions. Moreover, we use  $\mathcal{MC}$  to denote the universe of (finitely branching) DTMCs.

To derive a DTMC from an LTS, we have to calculate, for each states  $s$  and  $s'$  of LTS, the probability of moving from  $s$  to  $s'$ , by exploiting transition rates. Thus, we introduce two functions  $R : S \times S \mapsto \mathbb{R}_{\geq 0}$  and  $E : S \mapsto \mathbb{R}_{\geq 0}$ , such that

$$R(s, s') = \sum_{t \in TS(s, s')} \text{rate}(t) \quad \text{and} \quad E(s) = \sum_{s' \in S} R(s, s').$$

Intuitively,  $R(s, s')$  gives the rate of the set of transitions from  $s$  to  $s'$ , while  $E(s)$  computes the exit rate of states. The probability of moving from  $s$  to  $s'$  is derived from  $R(s, s')$  and  $E(s)$ , in the standard way.

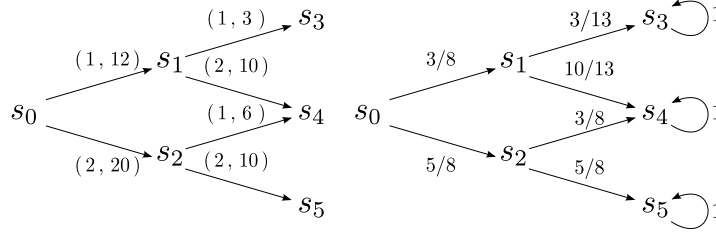
**Definition 3** (*Probabilistic Translation Function*). We define  $\mathcal{H} : \mathcal{LTS} \mapsto \mathcal{MC}$  as  $\mathcal{H}((S, s_0, \rightarrow)) = (S, s_0, P)$ , where  $P : S \mapsto Distr(S)$  is the probability transition function, s.t.,  $\forall s, s' \neq s \in S$  : if  $E(s) = 0$ , then  $P(s, s') = 0$ , and  $P(s, s) = 1$ ;  $P(s, s') = R(s, s')/E(s)$  otherwise.

Note that, traditionally, the semantics of a stochastic system is formalized as a *Continuous Time Markov Chain* (CTMC). We consider the DTMC because we are interested in probability of reachability properties (see the following section).

## 2.3. Probabilistic model checking

In the context of probabilistic model checking [35,34] we consider a fragment of the Probabilistic Computation Tree Logic (PCTL) [24], able to express probabilistic reachability properties. Probabilistic reachability captures the probability to reach a state which satisfies given property. Formally, this requires to evaluate the probability of a set of paths in the DTMC. We briefly recall main concepts concerning the validation of probabilistic reachability properties and we refer the interested reader to [35,3] for more details on PCTL model checking.

Let  $(S, s_0, P)$  be a DTMC. A *path*  $\pi$  is a non-empty (finite or infinite) ordered succession of states  $s_0, s_1, \dots$  of  $S$ . We denote the  $i^{\text{th}}$  state of the path  $\pi$  by  $\pi[i]$ , starting from 1, and the length of  $\pi$  by  $|\pi|$ , where  $|\pi| = \infty$  if  $\pi$  is infinite. The set of paths over  $S$  is denoted by  $Paths(S)$  and its subset of finite paths is denoted as  $FPaths(S)$ . For a finite path  $\pi$  we use  $\pi_{\text{last}}$  for the last state of the path. The *cylinder* corresponding to a path  $\pi$  is the set of all paths prefixed by  $\pi$ . Formally, for  $\pi \in Paths(S)$ ,  $C(\pi) = \{\pi\pi' \mid \pi' \in Paths(S)\}$  and  $C(s)$  denotes the set of paths starting from the state  $s$ .

Fig. 2.  $LTS(M_{ex})$ , and  $\mathcal{H}(LTS(M_{ex}))$ .

**Definition 4** (Probability of Paths). Let  $(S, s_0, P)$  be a DTMC. Let  $\Pi = \bigcup_{\pi \in FPaths(s)} C(\pi)$  be the set of all cylinders,  $\mathcal{B}$  be the smallest  $\sigma$ -algebra containing  $\Pi$ , and  $s \in S$  a state. The tuple  $(Paths(S), \mathcal{B}, P_s)$  is a probability space, where  $P_s$  is the unique measure satisfying, for all path  $s_0 \dots s_n$ ,

$$P_s(C(s_0 \dots s_n)) = \begin{cases} 1 & \text{if } s_0 = s \wedge n = 0 \\ P(s_0, s_1) \times \dots \times P(s_{n-1}, s_n) & \text{if } s_0 = s \wedge n > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Our reachability properties are parametric w.r.t. a set  $AP$  of propositional symbols (ranged over by  $\{A, B, \dots\}$ ). A symbol  $A \in AP$  denotes a set of conditions on multisets that are evaluated by a corresponding notion of satisfaction  $\models: \mathcal{S}(\Sigma) \times AP \mapsto \{\text{true}, \text{false}\}$ . As usual, given  $s \in \mathcal{S}(\Sigma)$  and  $A \in AP$ ,  $s \models A$  says that  $s$  satisfies  $A$ .

**Definition 5** (Concrete Reachability). Let  $mc = (S, s_0, P)$  be a DTMC. The probability of reaching a state satisfying  $A \in AP$ , starting from  $s \in S$ , is

$$Reach_{A,mc}(s) = P_s(\{\pi \in C(s) \mid \pi[i] \models A \text{ for some } i \geq 0\}).$$

Model checking of reachability properties on a DTMC, from a state  $s$ , consists of computing  $Reach_{A,mc}(s)$ , and can be done using standard iterative methods [35,3].

We use  $Reach(A)$  to denote  $Reach_{A,mc}(s_0)$  where  $mc = \mathcal{H}(LTS(M))$ , for a system  $M$  clear from the context.

**Example 1** (Concrete System Model Checking). We consider a simple system of chemical reactions where: starting from a configuration consisting of two molecules of  $X$ , two of  $Y$  and ten of  $W$ , two molecules  $X$  and  $Y$  may bind to form complex  $XY$ , or molecule  $X$  may be degraded by molecule  $W$ . Using  $\Sigma = \{X, Y, W, XY\}$ , the system can be formalized as  $M_{ex} = (\mathcal{R}, \mathcal{K}, s_0)$  where

$$\begin{aligned} s_0 &= \{(X, 2), (Y, 2), (W, 10), (XY, 0)\}, \\ \mathcal{K} &= \{k_1 = 3, k_2 = 1\}, \\ \mathcal{R} &= \{(\{X, Y\}, \{XY\}), (\{X, W\}, \{W\})\}. \end{aligned}$$

Note that we assume that the complexation is three times faster than the degradation. Fig. 2 shows the derived  $LTS$ ,  $LTS(M_{ex})$ , and the corresponding IMC,  $\mathcal{H}(LTS(M_{ex}))$ , where

$$\begin{aligned} S &= \{s_0 = \{(X, 2), (Y, 2), (W, 10), (XY, 0)\}, s_1 = \{(X, 1), (Y, 1), (W, 10), (XY, 1)\}, \\ s_2 &= \{(X, 1), (Y, 2), (W, 10), (XY, 0)\}, s_3 = \{(X, 0), (Y, 0), (W, 10), (XY, 2)\}, \\ s_4 &= \{(X, 0), (Y, 1), (W, 10), (XY, 1)\}, s_5 = \{(X, 0), (Y, 2), (W, 10), (XY, 0)\}\}. \end{aligned}$$

The probability of obtaining at least two complexes  $XY$  corresponds to the probability to reach  $s_3$ . That is,  $3/8 \times 3/13 = 9/104$ . This shows that, even if the rate of the complexation is (three times) greater than the one of the degradation, the concentration of reagent  $W$  makes the degradation more likely to happen than the binding of reagents  $X$  and  $Y$ .

### 3. Abstract systems modeling and model checking

We introduce *abstract systems*, the *abstract LTS semantics*, and the corresponding *abstract probabilistic semantics* in terms of IMC. Moreover, we prove the soundness of the approach, using notions of the abstract interpretation theory.

In order to approximate the information about the kinetic constants of the reaction rules we adopt the domain of *intervals of (non negative) reals*  $\mathbb{I}$  (the real valued version of intervals of integers [12,31,45]).

**Definition 6** (Intervals).  $\mathbb{I} = \{[m, n] \mid m \in \mathbb{R}_{\geq 0}, n \in \mathbb{R}_{\geq 0} \cup \{\infty\} \wedge m \leq n\}$ .

Over intervals of reals  $\mathbb{I}$  we use the operations and the order defined as follows.

$$\begin{aligned} \forall i, j \in \mathbb{I}, i &= [a, b], j = [c, d] : & [i]^- &= a, [i]^+ = b \\ i \times^{\mathbb{I}} j &= [a \times c, b \times d], & i \cup_{\mathbb{I}} j &= [\min(a, c), \max(b, d)], \\ i +^{\mathbb{I}} j &= [a + c, b + d], & i \sqsubseteq_{\mathbb{I}} j &\text{ iff } (i \cup_{\mathbb{I}} j = j). \end{aligned}$$

We consider both  $\cup_{\mathbb{I}}$  and  $\sqsubseteq_{\mathbb{I}}$  extended component-wise to  $m$ -sized vectors of intervals, and for simplicity we use the same symbols. For  $x \in \mathbb{R}_{\geq 0}$  we use  $x^\bullet = [x, x] \in \mathbb{I}$  for its *best abstraction* – i.e. the most precise abstraction – as interval, considered extended to vector of reals.

In *abstract systems* each reaction rule has associated an interval of reals ( $\mathcal{K}^\circ \in \mathbb{I}$ ) rather than a precise kinetic constant ( $k \in \mathbb{R}_{\geq 0}$ ).

**Definition 7** (*Abstract Systems*). An *abstract system*  $M$  is a triple  $(\mathcal{R}, \mathcal{K}^\circ, s_0)$  where the components  $\mathcal{R}$  and  $s_0$  are defined as in the concrete case, while  $\mathcal{K}^\circ = \{k_1^\circ, \dots, k_m^\circ\}$  is a *vector of intervals* with  $k_i^\circ \in \mathbb{I}$  for  $i \in \{1, \dots, m\}$ .

We denote the universe of abstract systems as  $\mathcal{M}^\circ$ . We assume the notations used for concrete systems extended, in the obvious way, to abstract systems.

The order  $\sqsubseteq_{\mathbb{I}}$  over intervals introduces a corresponding approximation order  $\sqsubseteq_{\mathcal{M}^\circ}$  over abstract systems.

**Definition 8** (*Order on Abstract Systems*). Let  $M_i^\circ \in \mathcal{M}^\circ$  for  $i \in \{1, 2\}$  be abstract systems. We say that  $M_1^\circ \sqsubseteq_{\mathcal{M}^\circ} M_2^\circ$  iff  $M_1^\circ \sim M_2^\circ \wedge \mathcal{K}(M_1^\circ) \sqsubseteq_{\mathbb{I}} \mathcal{K}(M_2^\circ)$ .

### 3.1. Abstraction and concretization

To formalize the relation between concrete and abstract systems we introduce a pair of functions, *abstraction* and *concretization functions*, which form a Galois connection [12]. The abstraction function  $\alpha$  reports the best approximation of sets of concrete systems differing only for the kinetic part of the rules: *sets of isomorphic systems*. Its counterpart is the concretization function which reports the set of concrete systems abstracted by an abstract system.

An abstract system represents infinite set of concrete systems differing only on the kinetic constants of reactions. To formalize this concept we introduce the domain of isomorphic concrete systems, sets of concrete systems which are identical except for the kinetic part of the rewriting rules.

Let  $\tilde{\mathcal{P}}(\mathcal{M}) = \{X \in \mathcal{P}(\mathcal{M}) \mid \forall M, M' \in X, M \sim M'\}$  be the domain of *sets of isomorphic concrete systems*. Given  $X \in \tilde{\mathcal{P}}(\mathcal{M})$  we denote with  $\mathcal{R}(X)$  and  $s_0(X)$  the shared components, e.g. the vector of rules and the starting state, respectively.

To define the concrete domain of the Galois connection we also have to define the order  $\sqsubseteq_{\tilde{\mathcal{P}}(\mathcal{M})}$  on sets of isomorphic concrete systems.

**Definition 9** (*Order on Set of Isomorphic Concrete Systems*). Let  $X_i \in \tilde{\mathcal{P}}(\mathcal{M})$  for  $i \in \{1, 2\}$  be sets of isomorphic concrete systems. We say that  $X_1 \sqsubseteq_{\tilde{\mathcal{P}}(\mathcal{M})} X_2$  iff  $\mathcal{K}_1 \sqsubseteq_{\mathbb{I}} \mathcal{K}_2$  where  $\mathcal{K}_i = \bigcup_{M \in X_i} (\mathcal{K}(M))^\bullet$ .

**Definition 10** (*Abstraction and Concretization Functions*). We define functions  $\alpha : \tilde{\mathcal{P}}(\mathcal{M}) \mapsto \mathcal{M}^\circ$  and  $\gamma : \mathcal{M}^\circ \mapsto \tilde{\mathcal{P}}(\mathcal{M})$  s.t.  $\forall X \in \tilde{\mathcal{P}}(\mathcal{M}), \forall M^\circ \in \mathcal{M}^\circ$  :

- $\alpha(X) = (\mathcal{R}(X), \overline{\mathcal{K}^\circ}, s_0(X))$  where  $\overline{\mathcal{K}^\circ} \equiv \bigcup_{M \in X} (\mathcal{K}(M))^\bullet$ ;
- $\gamma(M^\circ) = \{M \mid \alpha(\{M\}) \sqsubseteq_{\mathcal{M}^\circ} M^\circ\}$ .

This formalization shows that an abstract system  $M^\circ$  represents a (possibly infinite) set of isomorphic concrete systems  $\gamma(M^\circ)$ . Each model  $M \in \gamma(M^\circ)$  has the same vector of rules  $\mathcal{R}(M)$  and the same starting state  $s_0(M)$ , while the kinetic constants may vary in the vector of intervals  $\mathcal{K}^\circ(M^\circ)$ .

**Theorem 3.1.** *The pair of functions  $(\alpha, \gamma)$  is a Galois connection between  $(\tilde{\mathcal{P}}(\mathcal{M}), \sqsubseteq_{\tilde{\mathcal{P}}(\mathcal{M})})$  and  $(\mathcal{M}^\circ, \sqsubseteq_{\mathcal{M}^\circ})$ .*

**Proof.** We have to prove that: functions  $\alpha$  and  $\gamma$  are (i) *monotonic* and (ii) *adjoint*.

(i) Is trivial given the definition of  $\alpha$  and  $\gamma$ .

(ii) We have to show:  $\forall X \in \tilde{\mathcal{P}}(\mathcal{M}), M^\circ \in \mathcal{M}^\circ : \alpha(X) \sqsubseteq_{\mathcal{M}^\circ} M^\circ \Leftrightarrow X \sqsubseteq_{\tilde{\mathcal{P}}(\mathcal{M})} \gamma(M^\circ)$ .

Let us consider  $M^\circ = (\mathcal{R}, \mathcal{K}_{M^\circ}^\circ, s_0)$  and  $X = \{M_i = (\mathcal{R}', \mathcal{K}_i, s'_0), i \in I_X\}$ . By definition of  $\alpha$  and  $\gamma$ ,  $\gamma(M^\circ) = \{M_j = (\mathcal{R}, \mathcal{K}_j, s_0), j \in J_{\gamma(M^\circ)}\}$  and  $\alpha(X) = (\mathcal{R}', \overline{\mathcal{K}^\circ}, s'_0)$ , where  $\overline{\mathcal{K}^\circ} \equiv \bigcup_{i \in I_X} (\mathcal{K}(M_i))^\bullet$ . Thus, by definition of  $\sqsubseteq_{\tilde{\mathcal{P}}(\mathcal{M})}$  and  $\sqsubseteq_{\mathcal{M}^\circ}$ , it must be the case that  $\mathcal{R} = \mathcal{R}'$  and  $s_0 = s'_0$ , and, remains to show that

$$\overline{\mathcal{K}_X^\circ} \sqsubseteq_{\mathbb{I}} \mathcal{K}_{M^\circ}^\circ \Leftrightarrow \bigcup_{i \in I_X} (\mathcal{K}(M_i))^\bullet \sqsubseteq_{\mathbb{I}} \bigcup_{j \in J_{\gamma(M^\circ)}} (\mathcal{K}(M_j))^\bullet.$$

This is evident as the dis-equations are side by side equal by def. of  $\alpha$  and  $\gamma$ .  $\square$

### 3.2. Abstract LTS semantics

We introduce the LTS semantics associated with abstract systems, adopting an *abstract transition* relation  $s \xrightarrow{\eta, \beta^\circ}_\circ s'$ , where  $\eta$  is as in the concrete case, while  $\beta^\circ \in \mathbb{I}$  is an interval of rates.

The application of a rule  $R_\eta$  to a state  $s$  is modeled by the rule

$$\boxed{\frac{\mathcal{R}[\eta] = (l, r) \quad \mathcal{K}^\circ[\eta] = k^\circ \quad l \subseteq s' \quad \beta^\circ = \text{rate}^\circ(l, s, k^\circ)}{s \xrightarrow{\eta, \beta^\circ}_\circ s'} \quad (2)}$$

where  $\text{rate}^\circ(l, s, k^\circ) = \text{kin}(l, s) \times \mathbb{I} k^\circ$ . To compute  $\text{rate}^\circ(l, s, k^\circ)$  we follow the same reasoning of the concrete case, replacing exact rates with intervals.

We define the function  $LTS^\circ : \mathcal{M}^\circ \mapsto LTS^\circ$  such that  $LTS^\circ((\mathcal{R}, \mathcal{K}^\circ, s_0)) = (S, s_0, \rightarrow_\circ)$  is obtained by transitive closure of (2) starting from  $s_0$ . As in the concrete case the outgoing transitions from a state have distinct labels. In the following we use  $\mathcal{LTS}^\circ$  to denote the universe of abstract LTSs and we assume that the notations, defined for LTSs, are adapted in the obvious way to the abstract case.

To relate an LTS to its abstract counterpart (in particular to express the soundness and the precision of abstract LTSs) we introduce the concept of *best abstraction*, both for an LTS and for sets of isomorphic LTSs.

We say that two LTSs  $lts_i = (S_i, s_{0,i}, \rightarrow_i)$  for  $i \in \{1, 2\}$  are isomorphic ( $lts_1 \sim lts_2$ ) iff  $S_1 = S_2$  and  $s_{0,1} = s_{0,2}$ , and denote the universe of *isomorphic LTS* as  $\mathcal{P}(\mathcal{LTS})$ . Intuitively, two isomorphic LTS share the same state space, including the initial state.

**Definition 11** (*Best Abstraction of LTSs*). We define functions

- $\alpha_{\mathcal{LTS}^\circ} : \mathcal{LTS} \mapsto \mathcal{LTS}^\circ$  such that  $\alpha_{\mathcal{LTS}^\circ}((S, s_0, \rightarrow)) = ((S, s_0, \rightarrow_\circ^\alpha))$  with

$$\rightarrow_\circ^\alpha = \{s \xrightarrow{\eta, \beta^\bullet}_\circ s' \mid s \xrightarrow{\eta, \beta}_\circ s' \in \rightarrow\};$$

- $\hat{\alpha}_{\mathcal{LTS}^\circ} : \mathcal{P}(\mathcal{LTS}) \mapsto \mathcal{LTS}^\circ$  such that  $\hat{\alpha}_{\mathcal{LTS}^\circ}(X) \equiv (S(X), s_0(X), \rightarrow_\circ^\wedge)$  with

$$\rightarrow_\circ^\wedge = \left\{ s \xrightarrow{\eta, \hat{\beta}^\circ}_\circ s' \mid (S, s_0, \rightarrow) \in X_{LTS}, \hat{\beta}^\circ = \bigcup_{s \xrightarrow{\eta, \beta^\circ}_\circ s', (S, s_0, \rightarrow_\circ) \in X_{LTS}} \beta^\circ \right\}$$

where  $X_{LTS} = \{\alpha_{\mathcal{LTS}^\circ}(LTS(M)) \mid M \in X\}$ .

The most precise abstraction of an LTS is obviously obtained by replacing the rate  $\beta$  of each transition with the corresponding exact interval  $\beta^\bullet = [\beta, \beta]$ . Thus, function  $\alpha_{\mathcal{LTS}^\circ}$  does not effectively introduce any approximation. By contrast, the abstraction of a set of isomorphic LTSs, calculated by function  $\hat{\alpha}_{\mathcal{LTS}^\circ}$ , takes the rate which is the union of exact rates, corresponding to each concrete LTS.

Function  $\alpha_{\mathcal{LTS}^\circ}$  can be used to relate concrete and abstract LTS. To express soundness, however, we need to introduce an approximation order  $\sqsubseteq_{\mathcal{LTS}^\circ}$  over abstract LTSs, in the style of [13]. In this way, we can say that an abstract LTS  $lts^\circ \in \mathcal{LTS}^\circ$  is a *sound approximation* of an LTS  $lts \in \mathcal{LTS}$ , if it approximates the best abstraction of  $lts$ . That is  $\alpha_{\mathcal{LTS}^\circ}(lts) \sqsubseteq_{\mathcal{LTS}^\circ} lts^\circ$ .

**Definition 12** (*Order on Abstract LTSs*). Let  $lts_i^\circ = (S, s_0, \rightarrow_\circ^i)$  for  $i \in \{1, 2\}$  two abstract LTS. We say that  $lts_1^\circ \sqsubseteq_{\mathcal{LTS}^\circ} lts_2^\circ$  iff,  $\forall s, s' \in S$

$$\forall t_1^\circ = (s \xrightarrow{\eta, \beta_1^\circ}_\circ s') \in \rightarrow_\circ^1, \exists t_2^\circ = (s \xrightarrow{\eta, \beta_2^\circ}_\circ s') \in \rightarrow_\circ^2 \text{ such that } \beta_1^\circ \sqsubseteq \beta_2^\circ.$$

Intuitively,  $lts_1 \sqsubseteq_{\mathcal{LTS}^\circ} lts_2$  requires that, each couple of states, in  $\rightarrow_\circ^1$  relation, are in  $\rightarrow_\circ^2$  relation with a coarser transition rate interval.

Function  $\hat{\alpha}_{\mathcal{LTS}^\circ}$  can suitably be used to relate a set of isomorphic LTS with abstract LTS. More in details, the following theorem shows that  $LTS^\circ(M^\circ)$ , for an abstract system  $M^\circ$ , coincides with the best abstraction of the set of isomorphic LTS  $\{(LTS(M)) \mid M \in \gamma(M^\circ)\}$ . This demonstrates the precision of the abstract LTS semantics of an abstract system  $M^\circ$  with respect to the set of LTS describing the behavior of the concrete system  $M$  approximated by  $M^\circ$  (i.e.  $M \in \gamma(M^\circ)$ ).

**Theorem 3.2** (*Precision of  $LTS^\circ$* ). Let  $M^\circ \in \mathcal{M}^\circ$  be an abstract system. We have

$$\hat{\alpha}_{\mathcal{LTS}^\circ}(\{(LTS(M)) \mid M \in \gamma(M^\circ)\}) = LTS^\circ(M^\circ).$$



**Proof.** Let  $M^\circ = (\mathcal{R}, \mathcal{K}^\circ, s_0)$  and  $\widehat{LTS}(M^\circ) = \{LTS(M) \mid M \in \gamma(M^\circ)\}$ . Moreover, let  $LTS^\circ(M^\circ) = (S, s_0, \rightarrow_{\circ}^{M^\circ})$ . For each  $M \in \gamma(M^\circ)$  we have  $M = (\mathcal{R}, \mathcal{K}, s_0)$  for some vector of kinetic constants  $\mathcal{K}$ , and  $LTS(M) = (S, s_0, \rightarrow)$  for some transition relation  $\rightarrow$ ; consequently  $\widehat{\alpha}_{\mathcal{LTS}}(\widehat{LTS}(M^\circ)) = (S, s_0, \rightarrow_{\circ}^{\wedge})$  for some transition relation  $\rightarrow_{\circ}^{\wedge}$ . Hence, we have only to prove that  $\rightarrow_{\circ}^{\wedge} \Rightarrow \rightarrow_{\circ}^{M^\circ}$ .

Since the vector of rules  $\mathcal{R}$  in  $M^\circ$  is the same as in any  $M \in \gamma(M^\circ)$ , we have that each transition in  $LTS^\circ(M^\circ)$  has a corresponding transition in  $LTS(M)$ , namely

$\forall s, s' \in S, \forall t_1 = s \xrightarrow{\eta, \beta} s', \exists t_2 = s \xrightarrow{\eta, \beta^\circ} s'$ , and consequently,  $\forall t_1 = s \xrightarrow{\eta, \beta_1^\circ} s', \exists t_2 = s \xrightarrow{\eta, \beta_2^\circ} s'$ . Now, also  $\beta_1^\circ = \beta_2^\circ$  holds as, by def. of  $\bigcup^{\mathbb{I}}$ ,  $LTS$ , and  $\gamma$ ,

$$\begin{aligned} \beta_1^\circ &= \bigcup_{\substack{s \xrightarrow{\eta, \beta^\circ} s' \\ (S, s_0, \rightarrow_\circ) \in \alpha_{\mathcal{LTS}}(\widehat{LTS}(M^\circ))}} \beta^\circ = \left[ \min_{\substack{s \xrightarrow{\eta, \beta} s' \\ (S, s_0, \rightarrow) \in LTS(M^\circ)}} \beta, \max_{\substack{s \xrightarrow{\eta, \beta} s' \\ (S, s_0, \rightarrow) \in LTS(M^\circ)}} \beta \right] \\ &= \left[ \min_{(\mathcal{R}, \mathcal{K}, s_0) \in \gamma(M^\circ)} k_\eta \times \text{kin}(l_\eta, s), \max_{(\mathcal{R}, \mathcal{K}, s_0) \in \gamma(M^\circ)} k_\eta \times \text{kin}(l_\eta, s) \right] = k_\eta^\circ \times \text{kin}(l_\eta, s) \\ &= \beta_2^\circ. \quad \square \end{aligned}$$

A consequence of the previous Theorem is that  $LTS^\circ(M^\circ)$  is a sound approximation of  $LTS(M)$ , for each  $M$  represented by  $M^\circ$  (i.e. such that  $M \in \gamma(M^\circ)$ ).

**Corollary 3.3** (Soundness of  $LTS^\circ$ ). Let  $M^\circ \in \mathcal{M}^\circ$  be an abstract system and  $M \in \mathcal{M}$  such that  $M \in \gamma(M^\circ)$ . We have

$$\alpha_{\mathcal{LTS}}(LTS(M)) \sqsubseteq_{\mathcal{LTS}^\circ} LTS^\circ(M^\circ).$$

**Proof.** Follows from  $\alpha_{\mathcal{LTS}}(LTS(M)) \sqsubseteq_{\mathcal{LTS}^\circ} \widehat{\alpha}_{\mathcal{LTS}}(\{(LTS(M)) \mid M \in \gamma(M^\circ)\})$ .  $\square$

### 3.3. Interval Markov chains

We use *Interval Discrete-Time Markov Chains* [27,33] (IMC) to define the probabilistic semantics of abstract systems. We briefly recall the main concepts concerning the validation of probabilistic temporal properties on IMC and we refer the interested reader to [27,33,19].

**Definition 13** (Interval Markov Chain). An IMC is a tuple  $(S, s_0, P^-, P^+)$ , where:

- $S$  is the set of states and  $s_0 \in S$  the starting state;
- $P^-, P^+ : S \mapsto \text{PDistr}(S)$  are the lower and upper probability transition functions such that  $\forall s, s' \in S, P^-(s, s') \leq P^+(s, s')$  and  $\sum_{s'' \in S} P^-(s, s'') \leq 1 \leq \sum_{s'' \in S} P^+(s, s'')$ .

Here,  $P^-(s, s')$  and  $P^+(s, s')$  define an interval of probability that represents lower and upper bounds for the transition probabilities of moving from  $s$  to  $s'$ . In the following we use  $\mathcal{MC}^\circ$  to denote the universe of IMCs.

In an IMC, for any state  $s$ , there is a choice for an *admissible distribution* yielding the probabilities to reach successor states. A distribution  $\rho \in \text{Distr}(S)$  is admissible for an IMC  $mc^\circ = (S, s_0, P^-, P^+)$  and a state  $s \in S$  iff  $\forall s' \in S : P^-(s, s') \leq \rho(s') \leq P^+(s, s')$ . We use  $\text{ADistr}_{mc^\circ}(s)$  for denoting the admissible distributions for state  $s$  and IMC  $mc^\circ$ . As in *Markov Decision Processes* (MDP), the non-determinism is resolved by schedulers. The notion of path for IMCs is analogous to that presented for DTMCs, and therefore it is convenient to use the same notation.

**Definition 14** (Scheduler). Let  $mc^\circ = (S, s_0, P^-, P^+)$  be an IMC. A scheduler is a function  $\mathbb{S} : \text{FPaths}(S) \mapsto \text{ADistr}_{mc^\circ}(\pi_{\text{last}})$  for each path  $\pi \in \text{FPaths}(S)$ . We use  $\text{Adm}(mc^\circ)$  for the set of schedulers on  $mc^\circ$ .

Given a scheduler  $\mathbb{S} \in \text{Adm}(mc^\circ)$  a probability space over paths can be defined analogously as for DTMCs (see Definition 4). Thus,  $P_s^\mathbb{S}$  stands for the probability starting from the state  $s$  w.r.t. the scheduler  $\mathbb{S}$ .

On IMCs, probabilistic reachability properties gives lower and upper bounds, obtained considering the minimum and maximum probabilities w.r.t. all schedulers.

**Definition 15** (Abstract Reachability). Let  $mc^\circ = (S, s_0, P^-, P^+)$  be an IMC. The lower and upper bound of the probability of reaching a state satisfying a propositional symbol  $A \in AP$ , starting from  $s \in S$ , are defined as follows:

$$\begin{aligned} \text{Reach}_{A, mc^\circ}^\circ(s) &= \left[ \inf_{\mathbb{S} \in \text{Adm}(mc^\circ)} P_s^\mathbb{S}(\{\pi \in C(s) \mid \pi[i] \models A \text{ for some } i \geq 0\}), \right. \\ &\quad \left. \sup_{\mathbb{S} \in \text{Adm}(mc^\circ)} P_s^\mathbb{S}(\{\pi \in C(s) \mid \pi[i] \models A \text{ for some } i \geq 0\}) \right]. \end{aligned}$$

We introduce the concepts necessary to state the soundness and the precision of IMCs, similarly as for LTSs. To relate DTMCs to IMCs, their abstract counterparts, we introduce the *best abstraction* of a DTMC and of *sets of isomorphic DTMCs*.

Two IMCs  $mc_i = (S_i, s_{0,i}, P_i)$  for  $i \in \{1, 2\}$  are isomorphic ( $mc_1 \sim mc_2$ ) iff  $S_1 = S_2$  and  $s_{0,1} = s_{0,2}$ . We denote the universe of isomorphic DTMC with  $\tilde{\mathcal{P}}(\mathcal{MC})$ .

**Definition 16** (Best Abstraction of DTMCs). We define functions

- $\alpha_{\mathcal{MC}} : \mathcal{MC} \mapsto \mathcal{MC}^\circ$  such that  $\alpha_{\mathcal{MC}}((S, s_0, P, P)) = (S, s_0, P, P)$ ;
- $\hat{\alpha}_{\mathcal{MC}} : \tilde{\mathcal{P}}(\mathcal{MC}) \mapsto \mathcal{MC}^\circ$  such that  $\hat{\alpha}_{\mathcal{MC}}(X) = (S(X), s_0(X), P_\wedge^-, P_\wedge^+)$  where  $\forall s, s' \in S(X)$   

$$P_\wedge^+(s, s') = \sup_{(S, P, s_0) \in X} P(s, s'), \quad P_\wedge^-(s, s') = \inf_{(S, P, s_0) \in X} P(s, s').$$

As for LTS, function  $\alpha_{\mathcal{MC}}$  does not introduce any approximation. Actually, the intervals of probability introduced by  $\alpha_{\mathcal{MC}}$  are exact so that  $\forall A \in AP, mc \in \mathcal{MC} : [Reach_{A, mc}(s_0)]^\bullet = Reach_{A, \alpha_{\mathcal{MC}}(mc)}^\circ(s_0)$ . By contrast, function  $\hat{\alpha}_{\mathcal{MC}}$ , given a set of isomorphic DTMC, takes, for each pair of states in the shared state space, the minimum and the maximum probability with respect to all the corresponding exact value.

Moreover, we introduce an approximation order  $\sqsubseteq_{\mathcal{MC}^\circ}$  over IMC, which is defined similarly as in [10,16].

**Definition 17** (Order on IMCs). Let  $mc_i^\circ = (S, s_0, P_i^-, P_i^+)$  for  $i \in \{1, 2\}$  be two IMCs. We say that  $mc_1^\circ \sqsubseteq_{\mathcal{MC}^\circ} mc_2^\circ$  iff  $\forall s \in S, ADistr_{mc_1^\circ}(s) \subseteq ADistr_{mc_2^\circ}(s)$ .

Intuitively, we say that  $mc_1^\circ \sqsubseteq_{\mathcal{MC}^\circ} mc_2^\circ$  iff, for each state  $s \in S$ , the set of admissible distributions for  $s$  in  $mc_1^\circ$  is included in the set of admissible distributions of  $s$  in  $mc_2^\circ$ .

The following theorem states the soundness of the order on IMCs for probabilistic reachability. In particular,  $mc_1^\circ \sqsubseteq_{\mathcal{MC}^\circ} mc_2^\circ$  guarantees that the lower and upper bounds for probabilistic reachability obtained for  $mc_1$  are included in the ones obtained for  $mc_2$ .

**Theorem 3.4.** Let  $mc_i^\circ = (S, s_0, P_i^-, P_i^+)$  for  $i \in \{1, 2\}$ , be two IMCs. If  $mc_1^\circ \sqsubseteq_{\mathcal{MC}^\circ} mc_2^\circ$  then  $\forall A \in AP, s \in S : Reach_{A, mc_2^\circ}^\circ(s) \sqsubseteq_{\mathbb{I}} Reach_{A, mc_1^\circ}^\circ(s)$ .

**Proof.** We examine only the case  $[Reach_{A, mc_2^\circ}(s)]^- \leq [Reach_{A, mc_1^\circ}(s)]^-$ .

From  $mc_1^\circ \sqsubseteq_{\mathcal{MC}^\circ} mc_2^\circ$  it follows that  $ADistr_{mc_1^\circ}(s) \subseteq ADistr_{mc_2^\circ}(s)$ .

In order to simplify the proof it is convenient to exploit the fact that  $Reach_{A, mc^\circ}^-(s)$  can be specified as a linear equations system [10,16,35,19]. In particular, for  $h \in \{1, 2\}$ ,  $[Reach_{A, mc_h^\circ}(s)]^- = \bigcup_{i \in \{0, \infty\}} \rho_{A, mc_h^\circ}^{-,i}(s)$  where

$$\rho_{A, mc_h^\circ}^{-,i}(s) = \begin{cases} 1 & \text{if } s \models A, \\ 0 & \text{if } i = 0 \wedge s \not\models A, \\ \inf_{\rho_{j_h} \in ADistr_{mc_h^\circ}(s)} \sum_{s' \in S} \rho_{j_h}(s') \times \rho_{A, mc_h^\circ}^{-,i-1}(s') & \text{otherwise} \end{cases}$$

and where  $\bigcup$  stands for the least upper bound with respect to the underlying order on pseudo-distributions, e.g.  $\rho_1 \subseteq \rho_2$  iff for each  $s$ ,  $\rho_1(s) \leq \rho_2(s)$ .

Intuitively,  $\rho_{A, mc_h^\circ}^{-,i}(s)$  reports the minimum probability to reach a state satisfying  $A$ , starting from  $s$ , after  $i$ -iterates.

Therefore, it is enough to show that  $\rho_{A, mc_2^\circ}^{-,i}(s) \leq \rho_{A, mc_1^\circ}^{-,i}(s)$ , for every  $i \geq 0$ . The proof proceeds by induction.

( $i = 0$ ) There are two possibilities for  $\rho_{A, mc_2^\circ}^{-,0}(s)$ . Either  $s \models A$  and result is 1 or  $s \not\models A$  and the result is zero. Both the cases are trivial as  $\rho_{A, mc_2^\circ}^{-,0}(s) = \rho_{A, mc_1^\circ}^{-,0}(s)$ .

( $i > 0$ ) There are two possibilities for  $\rho_{A, mc_2^\circ}^{-,i}(s)$ . Either  $s \models A$  and result is 1 or the result is computed by

$$\rho_{A, mc_2^\circ}^{-,i}(s) = \inf_{\rho_{j_2} \in ADistr_{mc_2^\circ}(s)} \sum_{s' \in S} \rho_{j_2}(s') \times \rho_{A, mc_2^\circ}^{-,i-1}(s'). \quad (3)$$

The case of  $s \models A$  is trivial, as we have explained in the case of  $i = 0$ . In case (3), we observe that for  $\rho_{A, mc_1^\circ}^{-,i}(s)$  the result is

$$\rho_{A, mc_1^\circ}^{-,i}(s) = \inf_{\rho_{j_1} \in ADistr_{mc_1^\circ}(s)} \sum_{s' \in S} \rho_{j_1}(s') \times \rho_{A, mc_1^\circ}^{-,i-1}(s'). \quad (4)$$

In this case we have to compare (3) and (4). By inductive hypothesis we have that  $\rho_{A, mc_2^\circ}^{-,i-1}(s') \leq \rho_{A, mc_1^\circ}^{-,i-1}(s')$ , so that we reduce to show

$$\inf_{\rho_{j_2} \in ADistr_{mc_2^\circ}(s)} \sum_{s' \in S} \rho_{j_2}(s') \leq \inf_{\rho_{j_1} \in ADistr_{mc_1^\circ}(s)} \sum_{s' \in S} \rho_{j_1}(s').$$

This is guaranteed by the fact that  $ADistr_{mc_1^\circ}(s) \subseteq ADistr_{mc_2^\circ}(s)$ .  $\square$



### 3.4. Derivation of abstract probabilistic semantics

We define the abstract probabilistic translation function  $\mathcal{H}^\circ : \mathcal{LTS}^\circ \mapsto \mathcal{MC}^\circ$ . Moreover, we prove the soundness and precision of the abstract probabilistic semantics using the notions of Section 3.3.

The abstract LTS reports on transitions the number of the rule which is applied and the interval representing a possible range for its rate. From this information both lower and upper bounds for the probabilities of moving from a state to another can be calculated.

Following the guidelines of the derivation of the DTMC from the concrete LTS, we introduce functions  $R^\circ : S \times S \mapsto \mathbb{I}$  and  $E^\circ : S \mapsto \mathbb{I}$  s.t.  $\forall s, s' \in S$ ,

$$R^\circ(s, s') = \sum_{t \in TS(s, s')} rate^\circ(t) \quad \text{and} \quad E^\circ(s) = \sum_{s' \in S} R^\circ(s, s').$$

Intuitively,  $R^\circ(s, s')$  reports the interval of rates corresponding to the move from  $s$  to  $s'$ , while  $E^\circ(s)$  is the abstract exit rate.

For all states  $s$  and  $s' \in S$ , both lower and upper bounds of the probability of moving from  $s$  to  $s'$  can be determined by exploiting  $R^\circ(s, s')$  and  $E^\circ(s)$ . For these purposes we need to consider the *worst case* and *best case* scenario, respectively. That is, the transition to be maximized (minimized) takes as rate value its upper (lower) bound and all the others take their lower (upper) bound. This reasoning has to be properly combined with the special cases when  $[E^\circ(s)]^+ = 0$  (the state  $s$  is stable) or  $[E^\circ(s)]^- = 0$  (the state  $s$  is stable for some values of kinetic constant of some rules).

**Definition 18** (Abstract Probabilistic Translation Function). We define function  $\mathcal{H}^\circ : \mathcal{LTS}^\circ \mapsto \mathcal{MC}^\circ$  such that  $\mathcal{H}^\circ((S, s_0, \rightarrow^\circ)) = (S, s_0, P^-, P^+)$ , where  $P^-, P^+ : S \mapsto PDistr(S)$  are obtained, for each  $s, s' \in S, s \neq s'$ , as follows:

- if  $[E^\circ(s)]^+ = 0$ , then  $P^+(s, s') = P^-(s, s') = 0, P^+(s, s) = P^-(s, s) = 1$ ;
- if  $[E^\circ(s)]^+ > 0$ , then
  - (a) if  $[E^\circ(s)]^- = 0$ , then  $P^+(s, s) = 1, P^-(s, s) = 0$
  - (b) if  $[R^\circ(s, s')]^- = 0$ , then  $P^-(s, s') = 0$  else  
 $P^-(s, s') = [R^\circ(s, s')]^- / ([R^\circ(s, s')]^- + \sum_{s'' \in S, s'' \neq s'} [R^\circ(s, s'')]^+)$
  - (c) if  $[R^\circ(s, s')]^+ = 0$ , then  $P^+(s, s') = 0$  else  
 $P^+(s, s') = [R^\circ(s, s')]^+ / ([R^\circ(s, s')]^+ + \sum_{s'' \in S, s'' \neq s'} [R^\circ(s, s'')]^-)$ .

We remark that the IMC  $\mathcal{H}^\circ(LTS^\circ(M^\circ))$  has the same number of states of each DTMC  $\mathcal{H}(LTS(M))$  where  $M$  is a system represented by  $M^\circ$ .

The following theorem states that the approximation order over abstract LTSs is preserved by the translation to IMCs.

**Theorem 3.5.** Let  $lts_i^\circ = (S, s_0, \rightarrow_i^\circ)$  for  $i \in \{1, 2\}$  be abstract LTS s.t.  $lts_1^\circ \sqsubseteq_{\mathcal{LTS}^\circ} lts_2^\circ$ . We have  $\mathcal{H}^\circ(lts_1^\circ) \sqsubseteq_{\mathcal{MC}^\circ} \mathcal{H}^\circ(lts_2^\circ)$ .

**Proof.** From  $lts_1^\circ \sqsubseteq_{\mathcal{LTS}^\circ} lts_2^\circ$  we have that  $\forall s, s' \in S$

$$\forall t_1^\circ = (s \xrightarrow{\eta, \beta_1^\circ} s') \in \rightarrow_{\circ}^1, \exists t_2^\circ = (s \xrightarrow{\eta, \beta_2^\circ} s') \in \rightarrow_{\circ}^2 \quad \text{such that } \beta_1^\circ \sqsubseteq \beta_2^\circ. \quad (5)$$

We have to prove that (5)  $\Rightarrow ADistr_{\mathcal{H}^\circ(lts_1^\circ)}(s) \subseteq ADistr_{\mathcal{H}^\circ(lts_2^\circ)}(s)$ .

By Definition 18 of  $\mathcal{H}^\circ$ , for  $i \in \{1, 2\}$ ,  $\mathcal{H}^\circ(lts_i^\circ) = (S, s_0, P_i^-, P_i^+)$ . Moreover,  $ADistr_{\mathcal{H}^\circ(lts_i^\circ)}(s) = \rho_i$  s.t.  $\forall s' \in Next(s) : P_i^-(s, s') \leq \rho_i(s') \leq P_i^+(s, s')$ .

We have that  $P_i^+, P_i^-$  are defined, according to  $\mathcal{H}^\circ$ , maximizing and minimizing  $R^\circ(s, s')/E^\circ(s)$ . Namely, for the general case we have that,

$$\begin{aligned} P_i^-(s, s') &= [R^\circ(s, s')]^- / \left( [R^\circ(s, s')]^- + \sum_{s'' \in S, s'' \neq s'} [R^\circ(s, s'')]^+ \right) \\ &= \left[ \sum_{s \xrightarrow{\eta, \beta^\circ} s' \in \rightarrow_{\circ}^i} \beta^\circ \right]^+ / \left( \left[ \sum_{s \xrightarrow{\eta, \beta^\circ} s' \in \rightarrow_{\circ}^i} \beta^\circ \right]^+ + \sum_{s'' \in S, s'' \neq s'} \left[ \sum_{s \xrightarrow{\eta', \beta^\circ} s'' \in \rightarrow_{\circ}^i} \beta^\circ \right]^- \right). \end{aligned}$$

By (5), we have that  $P_1^-(s, s') \leq P_2^-(s, s')$  and, for the same reasoning on  $P_1^+, P_2^+(s, s') \geq P_2^+(s, s')$ . Similarly, for the special cases, when  $[E^\circ(s)]^+ = 0$  or  $[E^\circ(s')]^- = 0$ , by (5) we have that  $P_1^- = P_2^-$  and  $P_1^+ = P_2^+$ .

Thus  $\forall s \in S, ADistr_{\mathcal{H}^\circ(lts_1^\circ)}(s) \subseteq ADistr_{\mathcal{H}^\circ(lts_2^\circ)}(s)$ .  $\square$

The following theorems show the soundness and the precision of the IMC obtained by our approach with respect to the concrete probabilistic semantics. Specifically, we relate the IMC  $\mathcal{H}^\circ(LTS^\circ(M^\circ))$  with the DTMC  $\mathcal{H}(LTS(M))$  for each  $M$  represented by  $M^\circ$ . Following the same reasoning done for LTSs, we exploit the abstraction functions  $\alpha_{\mathcal{MC}}$  and  $\hat{\alpha}_{\mathcal{MC}}$ , reporting the best abstraction of a DTMC and of a set of isomorphic DTMC, respectively.

To prove the main theorem, we introduce the following lemma, stating that  $\alpha_{\mathcal{MC}} \circ \mathcal{H} = \mathcal{H}^\circ \circ \alpha_{\mathcal{LTS}}$ .

**Lemma 3.6.** Let  $M \in \mathcal{M}$  be a system. We have  $\alpha_{\mathcal{MC}}(\mathcal{H}(LTS(M))) = \mathcal{H}^\circ(\alpha_{\mathcal{LTS}}(LTS(M)))$ .

**Proof.** Let  $M = (\mathcal{R}, \mathcal{K}, s_0)$ ,  $LTS(M) = (S, s_0, \rightarrow)$ . We have  $\mathcal{H}(LTS(M)) = (S, s_0, P)$  and  $\alpha_{\mathcal{MC}}(\mathcal{H}(LTS(M))) = (S, s_0, P, P)$ . On the other hand we have  $\alpha_{\mathcal{LTS}}(LTS(M)) = (S, s_0, \rightarrow^\alpha)$  where

$$\rightarrow^\alpha = \{s' \xrightarrow{\eta, \beta^\bullet} s'' \mid s' \xrightarrow{\eta, \beta} s'' \in \rightarrow\} \text{ and } \mathcal{H}^\circ(\alpha_{\mathcal{LTS}}(LTS(M))) = (S, s_0, P, P). \quad \square$$

The following theorem shows that the IMC  $\mathcal{H}^\circ(LTS^\circ(M^\circ))$  coincides with the best abstraction, obtained by means of  $\hat{\alpha}_{\mathcal{MC}}$ , of the set of isomorphic DTMCs  $\{\mathcal{H}(LTS(M)) \mid M \in \gamma(M^\circ)\}$ . As a consequence,  $\mathcal{H}^\circ(LTS^\circ(M^\circ))$  is also a sound approximation of each DTMC  $\mathcal{H}(LTS(M))$  such that  $M \in \gamma(M^\circ)$  (as stated by Corollary 3.8).

**Theorem 3.7** (Precision of IMC). Let  $M^\circ \in \mathcal{M}^\circ$  be an abstract system. We have

$$\hat{\alpha}_{\mathcal{MC}}(\{\mathcal{H}(LTS(M)) \mid M \in \gamma(M^\circ)\}) = \mathcal{H}^\circ(LTS^\circ(M^\circ)).$$

**Proof.** Let  $\hat{\mathcal{H}}(M^\circ) = \{\mathcal{H}(LTS(M)) \mid M \in \gamma(M^\circ)\}$  and  $\widehat{LTS}(M^\circ) = \{LTS(M) \mid M \in \gamma(M^\circ)\}$ . By Theorem 3.2 it is enough to prove  $\hat{\alpha}_{\mathcal{MC}}(\hat{\mathcal{H}}(M^\circ)) = \mathcal{H}^\circ(\hat{\alpha}_{\mathcal{LTS}}(\widehat{LTS}(M^\circ)))$ . Here  $\hat{\alpha}_{\mathcal{MC}}(\hat{\mathcal{H}}(M^\circ)) = (S, s_0, P_\wedge^-, P_\wedge^+)$  and  $\mathcal{H}^\circ(\hat{\alpha}_{\mathcal{LTS}}(\widehat{LTS}(M^\circ))) = (S, s_0, P^-, P^+)$ .

We show that  $P_\wedge^+ = P^+$ ; the same reasoning applies to  $P_\wedge^- = P^-$ .

By definition of  $\hat{\alpha}_{\mathcal{MC}}$  and  $\mathcal{H}$ , we have for the general case that,  $\forall s, s' \in S$ ,

$$\begin{aligned} P_\wedge^+(s, s') &\equiv \max_{(S, P, s_0) \in \hat{\mathcal{H}}(M^\circ)} P(s, s') \\ &= \max_{(S, s_0, \rightarrow) \in \widehat{LTS}(M^\circ)} R(s, s') / \left( R(s, s') + \sum_{s'' \neq s'} R(s, s'') \right) \\ &= \max_{(S, s_0, \rightarrow) \in \widehat{LTS}(M^\circ)} \left( \sum_{\substack{\eta, \beta \\ (s \xrightarrow{\eta, \beta} s') \in \rightarrow}} \beta \right) / \left( \sum_{\substack{\eta, \beta \\ (s \xrightarrow{\eta, \beta} s') \in \rightarrow}} \beta + \sum_{\substack{\eta', \beta' \\ (s \xrightarrow{\eta', \beta'} s'') \in \rightarrow, s' \neq s''}} \beta' \right). \end{aligned} \quad (6)$$

Moreover, by definition of  $\mathcal{H}^\circ$ , we have for the general case that,  $\forall s, s' \in S$ ,

$$\begin{aligned} P^+(s, s') &\equiv [R^\circ(s, s')]^+ / \left( [R^\circ(s, s')]^+ + \sum_{s'' \in S, s'' \neq s'} [R^\circ(s, s'')]^- \right) \\ &= \left[ \sum_{\substack{\eta, \beta^\circ \\ s \xrightarrow{\eta, \beta^\circ} s'}} \beta^\circ \right]^+ / \left( \left[ \sum_{\substack{\eta, \beta^\circ \\ s \xrightarrow{\eta, \beta^\circ} s'}} \beta^\circ \right]^+ + \sum_{s'' \in S, s'' \neq s'} \left[ \sum_{\substack{\eta', \beta'^\circ \\ s \xrightarrow{\eta', \beta'^\circ} s''}} \beta'^\circ \right]^- \right) \\ &= \sum_{\substack{\eta, \beta^\circ \\ s \xrightarrow{\eta, \beta^\circ} s'}} [\beta^\circ]^+ / \left( \sum_{\substack{\eta, \beta^\circ \\ s \xrightarrow{\eta, \beta^\circ} s'}} [\beta^\circ]^+ + \sum_{\substack{\eta', \beta'^\circ \\ s \xrightarrow{\eta', \beta'^\circ} s'', s' \neq s''}} [\beta'^\circ]^- \right). \end{aligned} \quad (7)$$

For the general case, it remains to prove that (6) = (7), that is true by the fact that maximizing  $a/(a+b)$  corresponds to maximize  $a$  and minimize  $b$ , and by definition of  $\gamma$ , which ensures that the maximum in  $\widehat{LTS}(M^\circ)$  of  $\sum_{\substack{\eta, \beta \\ (s \xrightarrow{\eta, \beta} s') \in \rightarrow}} \beta$  is equal to  $\sum_{\substack{\eta, \beta^\circ \\ s \xrightarrow{\eta, \beta^\circ} s'}} [\beta^\circ]^+$  and the minimum  $\widehat{LTS}(M^\circ)$  of  $\sum_{\substack{\eta', \beta' \\ (s \xrightarrow{\eta', \beta'} s'') \in \rightarrow, s' \neq s''}} \beta$  is equal to  $\sum_{s'' \in S, s'' \neq s'} \left[ \sum_{\substack{\eta', \beta'^\circ \\ s \xrightarrow{\eta', \beta'^\circ} s''}} \beta'^\circ \right]^-$ .

The special cases in which either  $P^+(s, s') = 1$  and  $s = s'$  or  $P^+(s, s') = 0$  and  $s \neq s'$  are trivial.  $\square$

**Corollary 3.8** (Soundness of IMC). Let  $M^\circ \in \mathcal{M}^\circ$  be an abstract system and  $M \in \mathcal{M}$  such that  $M \in \gamma(M^\circ)$ . We have

$$\alpha_{\mathcal{MC}}(\mathcal{H}(LTS(M))) \sqsubseteq_{\mathcal{MC}^\circ} \mathcal{H}^\circ(LTS^\circ(M^\circ)).$$

**Proof.** Follows from  $\alpha_{\mathcal{MC}}(\mathcal{H}(LTS(M))) \sqsubseteq_{\mathcal{MC}^\circ} \hat{\alpha}_{\mathcal{MC}}(\{\mathcal{H}(LTS(M)) \mid M \in \gamma(M^\circ)\})$ .  $\square$

The following theorem states our main result: the soundness and precision results on IMC are lifted to *probabilistic reachability*. More in details, the lower and upper bounds calculated over the IMC  $\mathcal{H}^\circ(LTS^\circ(M^\circ))$  for probabilistic reachability are exactly the *most precise values* which are correct. Indeed, they correspond to the minimum and the maximum of the exact probabilities for probabilistic reachability, calculated over the DTMC  $\mathcal{H}(LTS(M))$  for each concrete system  $M$  represented by  $M^\circ$ .

**Theorem 3.9.** Let  $M^\circ \in \mathcal{M}^\circ$  be an abstract system and  $mc^\circ = \mathcal{H}^\circ(LTS^\circ(M^\circ)) = (S, s_0, P^-, P^+)$  be the associated IMC. For each state  $s \in S$  and for each  $A \in AP$  we have

$$\bigcup_{mc \in \{\mathcal{H}(LTS(M)) \mid M \in \gamma(M^\circ)\}} [Reach_{A, mc}(s)]^\bullet = Reach_{A, mc^\circ}^\circ(s).$$

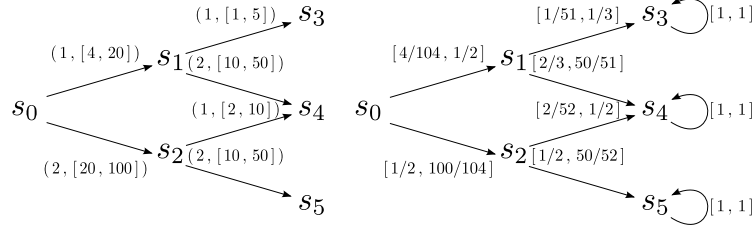


Fig. 3.  $LTS^o(M_{ex}^o)$  and  $H^o(LTS^o(M_{ex}^o))$ .

**Proof** (Sketch of proof). In the definition of  $Reach_{A,mc^o}^o(s)$  the probability of a path is computed by associating each step from a state  $s$  to a state  $s'$  in the path with a probability taken from one of the admissible distributions for  $s$  and  $mc^o$ . The proof reduces to show that for each state  $s$  of both  $\hat{\alpha}_{MC}^o$  and  $H^o(LTS^o(M^o))$ , we have that

$$(\exists M \in \gamma(M^o) \text{ s.t. } (H(LTS(M))) = (S, s_0, P) \wedge P(s) = \rho) \Leftrightarrow \rho \in ADistr_{H^o(LTS^o(M^o))}(s).$$

The implication  $\Rightarrow$  follows from  $M \in \gamma(M^o)$  and Theorem 3.7. The implication  $\Leftarrow$  follows from the definition of  $\gamma$ ,  $\gamma(M^o)$  contains a concrete model  $M$  for each possible combination of values chosen from the intervals of  $M^o$ , and from the fact that  $LTS^o$  and  $H^o$  do not introduce admissible distributions that are not present in any  $H(LTS(M))$ .  $\square$

Finally, we conclude that the IMC, derived from the abstract LTS of an abstract system  $M^o$ , gives conservative bounds for probability of reachability properties for each concrete system  $M \in \gamma(M^o)$ .

**Theorem 3.10.** Let  $M^o \in \mathcal{M}^o$  be an abstract system and  $mc^o = H^o(LTS^o(M^o)) = (S, s_0, P^-, P^+)$  be the associated IMC. For each  $M \in \mathcal{M}$ , such that  $M \in \gamma(M^o)$ , for each state  $s \in S$  and for each  $A \in AP$  we have,

$$[Reach_{A,H(LTS(M))}(s)]^\bullet \sqsubseteq_{\sqsubseteq} Reach_{A,mc^o}^o(s).$$

**Proof.** From Theorems 3.2 and 3.5, we obtain  $H^o(\alpha_{LTS^o}(\mathcal{LTS}(M))) \sqsubseteq_{MC^o} H^o(\mathcal{LTS}(M^o))$ . From Lemma 3.6,  $\alpha_{MC}(\mathcal{H}(\mathcal{LTS}(M))) \sqsubseteq_{MC^o} H^o(\mathcal{LTS}(M^o))$ . By Theorem 3.4,  $Reach_{A,\alpha_{MC}(\mathcal{H}(\mathcal{LTS}(M)))}(s) \sqsubseteq_{\sqsubseteq} Reach_{A,H^o(\mathcal{LTS}(M^o))}(s)$  and finally, by Definition 16,  $[Reach_{A,H(LTS(M))}(s)]^\bullet \sqsubseteq_{\sqsubseteq} Reach_{A,H^o(\mathcal{LTS}(M^o))}(s)$ .  $\square$

**Example 2** (Abstract System Model Checking). We consider the system of reactions introduced in Example 1. In this case, we assume that the kinetic constants of the rules are not exact, but described by intervals. For instance, we consider the abstract system  $M_{ex}^o = (\mathcal{R}, \mathcal{K}^o, s_0)$  where  $\mathcal{R}$  and  $s_0$  are the same as Example 1, while  $\mathcal{K}^o = \{k_1^o = [1, 5], k_2^o = [1, 5]\}$ . Note that the concrete system  $M_{ex}$  of Example 1 is one of the systems represented by  $M_{ex}^o$ , i.e.  $M_{ex} \in \gamma(M_{ex}^o)$ .

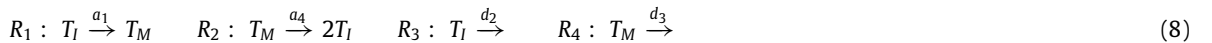
Fig. 3 shows the abstract LTS  $LTS^o(M_{ex}^o)$  and the corresponding IMC  $H^o(LTS^o(M_{ex}^o))$ , where the state space  $S$  is the same of Example 1. Thus, the probability of obtaining at least two complexes  $XY$  is again the probability to reach state  $s_3$ . In the abstract case, we obtain the interval of probability  $[4/104, 1/2] \times [1/51, 1/3] = [1/1326, 1/6]$ . This shows that the abstraction is precise enough to observe essentially the same behavior of Example 1 even if the reaction rates are imprecise: the concentration of reagent  $W$  makes the degradation more likely to happen than the binding of reagent  $X$  and  $Y$ .

#### 4. Case study: tumor cell growth

We briefly present the application of the proposed approach to a model of tumor growth, proposed by Villasana and Radunskaya and studied with Delay Differential Equations (DDEs) in [43].

Tumor growth is based on cell divisions (or *mitosis*). The cell cycle is the process between two mitosis and it consists of four phases: the  $G_1$  phase (a resting phase or gap period), the  $S$  phase where the replication of DNA occurs, the  $G_2$  gap period, and the mitosis phase  $M$  in which the cells segregate the duplicated sets of chromosomes between daughter cells. The three phases  $G_1$ ,  $S$ , and  $G_2$  constitute the pre-mitotic phase, also called *interphase*.

The simplest model proposed in [43] considers two populations of tumor cells: the population of tumor cells during cell cycle interphase, and the population of tumor cells during mitosis. Such a model can be expressed as the following reactions:



where  $T_I$  and  $T_M$  are tumor cells in interphase and in mitosis, respectively. Reaction  $R_1$  represents the passage of a tumor cell from the interphase to the mitosis phase,  $R_2$  represents the mitosis, whereas  $R_3$  and  $R_4$  represent tumor cell death.

Let  $d$  be the rate at which mitotic cells disappear, namely  $d = d_3 + a_4$ . Fig. 4 shows the results of the analytical study of the DDEs model done in [43], by setting the parameters  $a_4$  and  $d_2$  to 0.5 and 0.3, respectively, and by varying  $a_1$  and  $d$ . There are two regions. The region in which the tumor grows is R-I, while in R-II both kinds of tumor cells disappear. Note that since  $a_4 = 0.5$ , the area in the figure in which  $d < 0.5$  corresponds to negative values of  $d_3$  that are not realistic.

We have constructed three abstract systems modeling tumor growth  $M_1^o$ ,  $M_2^o$  and  $M_3^o$  by replacing kinetic constants in the reactions (8) with intervals. Actually, in all the three systems we have replaced  $a_1$  with  $[0.8, 0.9]$ ,  $a_4$  with  $0.5^\bullet$  and  $d_2$  with  $0.3^\bullet$ . As regards  $d_3$ , we have replaced it with  $[0.05, 0.1]$ ,  $[1, 1.4]$  and  $[0.005, 2]$  in  $M_1^o$ ,  $M_2^o$  and  $M_3^o$ , respectively.

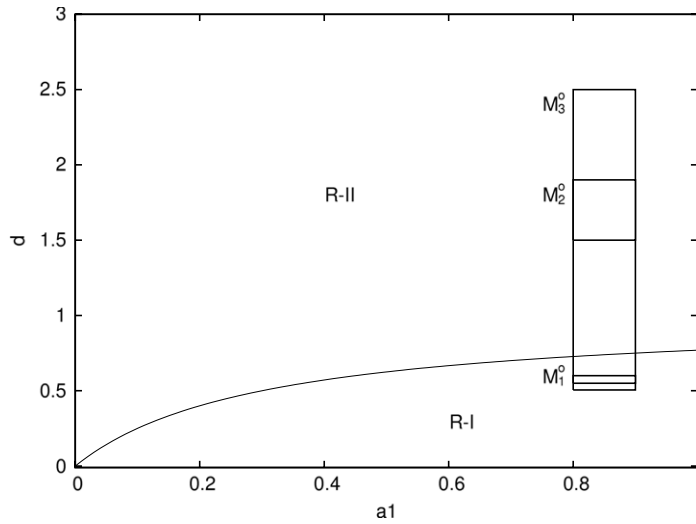


Fig. 4. The regions which describe the different behaviors of the DDEs model by varying parameters  $a_1$  and  $d$ .

This corresponds to consider a region in R-I, a region in R-II and a region across the line separating R-I and R-II (see Fig. 4). Moreover, we have considered an initial population consisting of 10 tumor cells in interphase and 10 tumor cells in mitosis.

Formally,  $M_i^o = (\mathcal{R}, \mathcal{K}_i^o, s_0)$  with  $i \in \{1, 2, 3\}$  where  $s_0 = \{(T_I, 10), (T_M, 10)\}$ ,

$$\mathcal{R} = \{(\{T_I\}, \{T_M\}), (\{T_M\}, \{2T_I\}), (\{T_I\}, \emptyset), (\{T_M\}, \emptyset)\}$$

$$\mathcal{K}_i^o = \{[0.8, 0.9], 0, 5^\bullet, 0, 3^\bullet, d_3^i\}$$

where  $d_3^1 = [0.05, 0.1]$ ,  $d_3^2 = [1, 1.4]$ ,  $d_3^3 = [0.005, 2]$ .

In order to perform model checking on the abstract systems we have developed a translator [1] of the abstract MSR semantics into MDP by following the extreme distribution approach of [19]. In particular, the translator computes in effective way the extreme distributions from the probability intervals reported by the IMC probabilistic semantics of the system. The tool invokes PRISM [34] for the verification of the properties on the corresponding MDP model.

In order to obtain a finite abstract probabilistic model we have heuristically limited the number of states of the model to  $10^4$ . Specifically, the abstract model is built iteratively by performing breadth-first visit of the abstract state space. We generate all the states having the number of individuals of both species less or equal to  $10^2$  and we introduce a special abstract state which represents all the other states. Such a special state has a self loop (that is the probability to move into any other state is zero). Moreover, the moves from states contained in the model to states represented by the special state, are replaced by corresponding moves to the special state.

Note that the technique applied is not correct in general given that it introduces an error in the probability of reachability properties. Intuitively, the error is significant if one wants to calculate the probability to reach a state which is represented by the special state. In this case, we are interested to calculate the probability to reach a state containing at most 30 occurrences of  $T_M$ . Thus, the effect of the finite truncation to a finite model is minimal and it corresponds to the probability of the paths which lead to the state to be reached and includes a state represented by the special state. The error introduced is over-approximated in the finite abstract model by the probability to reach the special state. Such a probability is negligible when the number of individuals of the species used in the truncation is sufficiently high.

In Fig. 5 we show the results of model checking of property  $\text{Reach}^o(T_M = x)$  in  $M_1^o$ ,  $M_2^o$  and  $M_3^o$  by varying  $x$ . In  $M_1^o$  both the minimum and the maximum probabilities tend to zero for small values of  $x$  while they are both equal to 1 for values greater than or equal to 10 (the initial value of  $T_M$ ). In  $M_2^o$  the opposite holds. In  $M_3^o$  we have that both probabilities are equal to 1 when  $x$  is 10, but they tend to the interval  $[0, 1]$ , namely to complete uncertainty, both for greater and smaller values of  $x$ . A more immediate representation of the dynamic behavior of the considered systems is the plot of  $\text{Reach}(T_M = y \wedge \text{time} = x)$  given in Fig. 6, where  $\text{time}$  is the number of steps in the path that reaches a state satisfying  $T_M = y$ . Note that results would not change significantly by varying the size of the three areas. What really matters in this case study is whether parameters are taken from region R-I or R-II.

The obtained results agree with analytic ones: the results on  $M_1^o$  describe tumor growth, those on  $M_2^o$  describe tumor decay, those on  $M_3^o$  leave uncertainty.

Our approach is more precise with respect to analytic studies, as it looks at all possible behaviors of the modeled system, rather than a single average behavior. Moreover, a more realistic discrete probabilistic semantics is considered, instead of a continuous deterministic one.

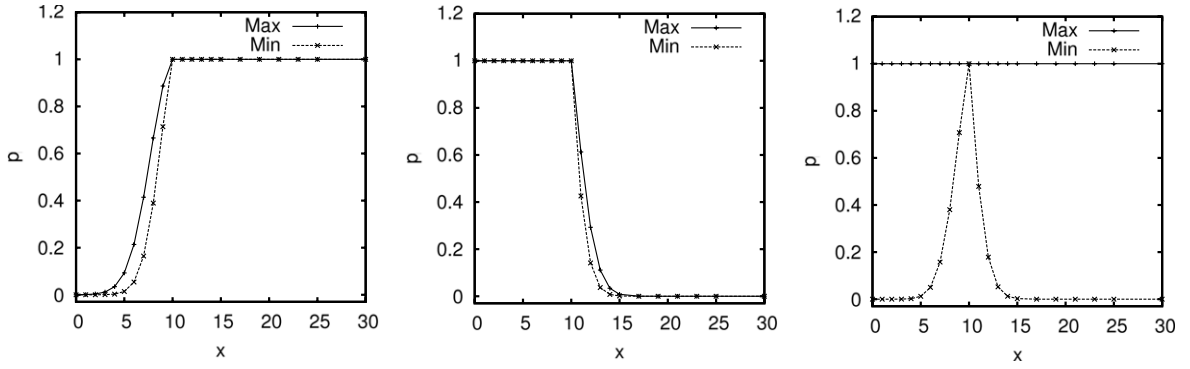


Fig. 5. Model checking of  $\text{Reach}^o(T_M = x)$  in, from left to right,  $M_1^o, M_2^o, M_3^o$ .

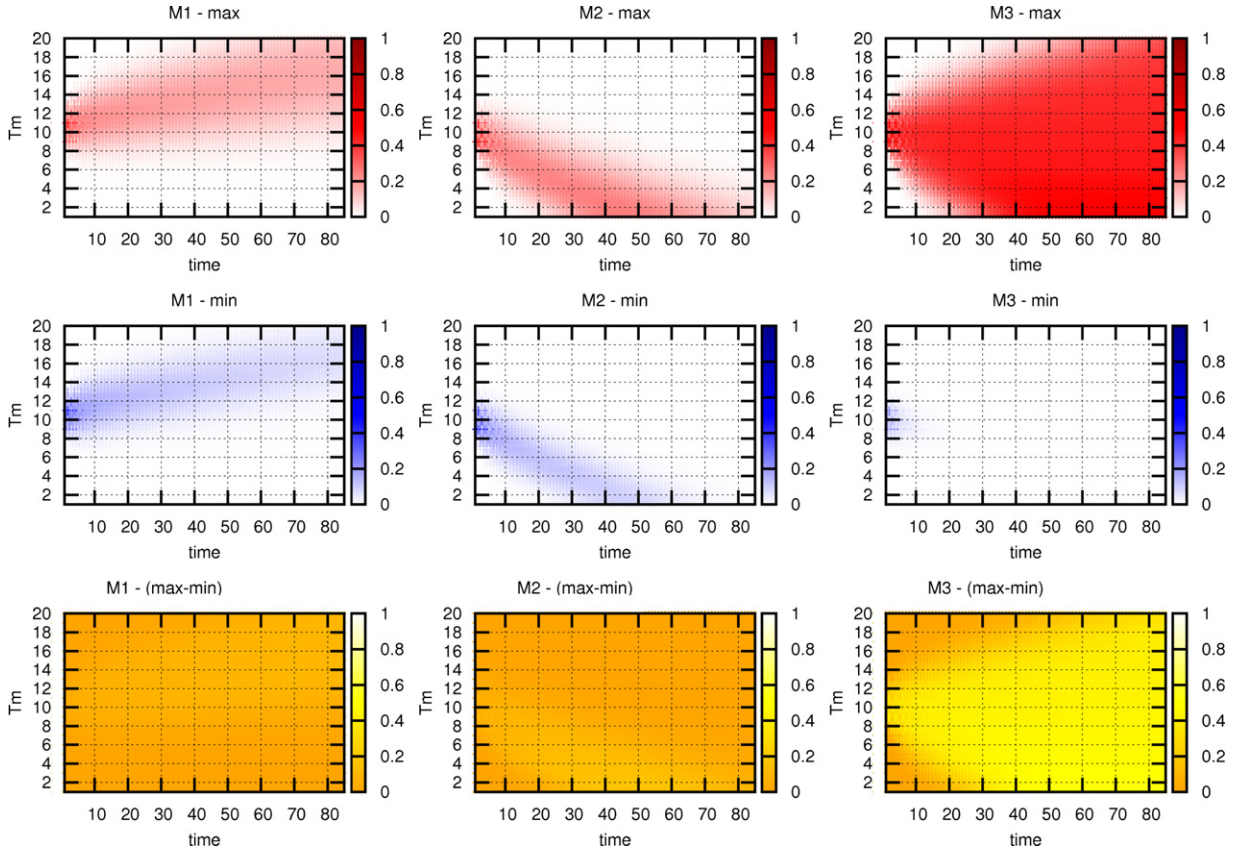


Fig. 6. Model checking of  $\text{Reach}(T_M = y \wedge \text{time} = x)$  in, from left to right,  $M_1^o, M_2^o, M_3^o$ . Probabilities expressed by color intensity. The plots of  $[\text{Reach}]^+$ ,  $[\text{Reach}]^-$  and  $[\text{Reach}]^+ - [\text{Reach}]^-$  are shown from top to bottom.

## 5. Related works

The design of abstractions for probabilistic or stochastic models has been widely investigated over the last few years.

Most of the proposals study abstractions able to deal with the traditional state-explosion problem, which limits the practical application of probabilistic model checking. The proposals in [19,16,40,26,41,42] present similar approaches for approximating probabilistic models, based on MDP and IMC. In these approaches the abstract model is derived from the concrete one by considering a partition of the concrete state space, and by computing, for each abstract state, the abstract probabilities from the concrete probabilities. The technique proposed in [36] extends the approach of [16] in order to better approximate non-deterministic and probabilistic systems modeled by MDP. The abstract model is based on two-player stochastic games that are able to separate the non-determinism introduced by the abstraction from the non-determinism present in the concrete MDP. Katoen et al. [28] proposes an extension of the approximation technique presented in [19] to



Continuous-Time Markov Chains (CTMC). The approach uses uniform CTMC [2] where abstract transition probabilities are approximated by means of intervals.

The proposals of [30,44] investigate the implementation of the abstraction techniques for MDP, proposed in [16,30]. The methodology applies predicate abstraction to PRISM models and supports the effective construction of an abstract model, using an extension of the PRISM-language. Other approaches for handling the state-explosion problem are those based on infinite state abstraction [23], on symmetry reduction [18] or counter-example driven abstraction refinement [25]. We refer to [29,32] for a more detailed discussion of abstraction techniques for probabilistic models.

We use abstraction techniques in a different way in order to deal with the uncertainty about kinetic rates, typical of biological system modeling. In our context the abstract probabilistic model (IMC) represents an infinite set of concrete models with different kinetic rates and is calculated in effective way from an LTS semantics. A similar approach is presented in [10,20] to validate probabilistic reachability properties of biological systems, modeled in the stochastic process algebra Chemical Ground Form [7]. The analysis is based on the idea of representing a set of experiments, which differ only for the initial concentration of reagents by using intervals. The abstract probabilistic semantics modeled as an IMC is systematically derived from an abstract LTS.

The methodology presented in [14,15] applies abstract interpretation techniques to biological systems models, specifically signaling pathways. The proposed analysis calculates information about the reachable complexes, which could be generated at run-time, and permits to generate smaller systems of differential equations from the concrete ones.

Finally, [38,17] investigate the application of abstract interpretation in the context of standard concurrent probabilistic programming languages.

## 6. Conclusions

In this paper we have considered biological systems modeled by MSR, where rewriting rules, corresponding to reactions, are enriched by real valued kinetic constants. Our framework based on abstract interpretation supports probabilistic model checking of MSR systems with uncertain kinetic rates. Model checking an abstract system gives conservative probabilistic bounds with respect to the (infinite) set of concrete models which are abstracted. The abstract probabilistic semantics (IMC) is derived in a systematic way from an abstract LTS semantics. This approach allows us to safely and effectively manage in a finite way the semantics of an infinite set of (finite) systems. Moreover results obtained on an abstract system are exactly the most precise values which are correct. Indeed, they correspond to the minimum and the maximum of the concrete probability values corresponding to each concrete system represented by an abstract one.

We have developed an automatic verifier of abstract systems [1]: the tool, based on PRISM [34], uses a translation (similar to that of [19]) of the abstract probabilistic model (IMC) into a MDP. This translation has exponential complexity as it requires the computation of all the extreme distributions whose number grows exponentially with the number of uncertain parameters. More efficient algorithms, which calculate the extreme distributions on-the-fly, could be applied (see [19,40]). The approximated verification approach presented by [21] could also be used.

We have applied the proposed approach to a model of tumor growth [43], obtaining more precise results than the ones given by analytical studies.

As regards future development of our work, since in the presented case study we made use of an ad-hoc techniques to deal with infinite state space, it would be interesting to combine the proposed approach, dealing with uncertainty of kinetic rates, with abstraction approaches dealing with systems with infinite state space [6,9,37,22].

Moreover, we plan to investigate the extension of our methodology to the abstraction of CTMC, for example by following the approach of [28], based on uniform CTMC. In particular, from transition rates of a LTS it might be possible to derive a uniform CTMC. The idea is to consider the same exit rate for all states, which should be a value greater than the exit rate of any state in the LTS. The new exit rate can then be used in all CTMC states to compute probability distributions in place of the exit rate for that state obtained from the LTS. A abstract uniform CTMC could be obtained by choosing as exit rate the greatest of all exit rates of the uniform CTMCs to be abstracted.

## References

- [1] AMSR2PRISM tool: <http://www.di.unipi.it/msvbio/wiki/amsr2prism>.
- [2] C. Baier, H. Hermanns, J.P. Katoen, B.R. Haverkort, Efficient computation of time-bounded reachability probabilities in uniform continuous-time Markov decision processes, *Theoret. Comput. Sci.* 345 (2005) 2–26.
- [3] C. Baier, J.P. Katoen, *Principles of Model Checking*, The MIT Press, 2008.
- [4] R. Barbuti, G. Caravagna, A. Maggiolo-Schettini, P. Milazzo, G. Pardini, The calculus of looping sequences, in: M. Bernardo, P. Degano, G. Zavattaro (Eds.), *Formal Methods for Computational Systems Biology (SFM 2008)*, in: LNCS, vol. 5016, Springer, 2008, pp. 387–423.
- [5] R. Barbuti, F. Levi, P. Milazzo, G. Scatena, Probabilistic model checking of biological systems with uncertain kinetic rates, in: O. Bournez, I. Potapov (Eds.), *Reachability Problems, RP 2009*, in: LNCS, vol. 5797, Springer, 2009, pp. 64–78.
- [6] S. Bensalem, Y. Lakhnech, S. Owre, Computing abstractions of infinite state systems compositionally and automatically, in: A.J. Hu, M.Y. Vardi (Eds.), *Computer Aided Verification, CAV 1998*, in: LNCS, vol. 1427, Springer, 1998, pp. 319–331.
- [7] L. Cardelli, On process rate semantics, *Theoret. Comput. Sci.* 391 (2008) 190–215.
- [8] I. Cervesato, N.A. Durgin, P. Lincoln, J.C. Mitchell, A. Scedrov, A meta-notation for protocol analysis, in: *IEEE Computer Security Foundations Workshop, CSFW 1999*, IEEE Computer Society, 1999, pp. 55–69.
- [9] A. Chutinan, B.H. Krogh, Verification of infinite-state dynamic systems using approximate quotient transition systems, *Trans. Autom. Control* 26 (2001) 1401–1410.



- [10] A. Coletta, R. Gori, F. Levi, Approximating probabilistic behaviors of biological systems using abstract interpretation, in: *Biology to Concurrency and Back, FBTC 2008*, in: ENTCS, vol. 229, Elsevier, 2009, pp. 165–182.
- [11] M. Coppo, F. Damiani, M. Drocco, E. Grassi, A. Troina, Stochastic calculus of wrapped compartments, in: A. Di Pierro, G. Norman (Eds.), *Quantitative Aspects of Programming Languages, QAPL 2010*, in: EPTCS, vol. 28, 2010, pp. 82–98.
- [12] P. Cousot, R. Cousot, Abstract interpretation: a unified lattice model for static analysis of programs by construction or approximation of fixpoints, in: *ACM Symposium on Principles of Programming Languages, POPL 1977*, ACM Press, 1977, pp. 238–252.
- [13] D. Dams, R. Gerth, O. Grumberg, Abstract interpretation of reactive systems, *TOPLAS* 19 (1997) 253–291.
- [14] V. Danos, J. Feret, W. Fontana, J. Krivine, Abstract interpretation of cellular signalling networks, in: F. Logozzo, D. Peled, L.D. Zuck (Eds.), *Verification, Model Checking, and Abstract Interpretation, VMCAI 2008*, in: LNCS, vol. 4905, Springer, 2008, pp. 83–97.
- [15] V. Danos, J. Feret, W. Fontana, H. Russell, J. Krivine, Abstracting the differential semantics of rule-based models: exact and automated model reduction, in: *Annual IEEE Symposium on Logic in Computer Science, LICS 2010*, IEEE Computer Society, 2010, pp. 362–381.
- [16] P.R. D'Argenio, B. Jeannet, H.E. Jensen, K.G. Larsen, Reachability analysis of probabilistic systems by successive refinements, in: L. de Alfaro, S. Gilmore (Eds.), *Process Algebra and Probabilistic Methods, Performance Modeling and Verification, PAPM-PROBMIV*, in: LNCS, vol. 2165, Springer, 2001, pp. 39–56.
- [17] A. Di Pierro, H. Wiklicky, Concurrent constraint programming: towards probabilistic abstract interpretation, in: *Principles and Practice of Declarative Programming, PPDP 2000*, ACM Press, 2000, pp. 127–138.
- [18] A. Donaldson, A. Miller, D. Parker, GRIP: generic representatives in PRISM, in: *Fourth International Conference on the Quantitative Evaluation of Systems, QEST 2007*, IEEE Computer Society, 2007, pp. 115–116.
- [19] H. Fecher, M. Leucker, V. Wolf, Don't know in probabilistic systems, in: Antti Valmari (Ed.), *Model Checking Software, 13th International SPIN Workshop*, in: LNCS, vol. 3925, Springer, 2006, pp. 71–88.
- [20] R. Gori, F. Levi, Abstract interpretation for probabilistic termination of biological systems, in: G. Ciobanu (Ed.), *Membrane Computing and Biologically Inspired Process Calculi, MeCBIC 2009*, in: EPTCS, vol. 11, 2009, pp. 137–153.
- [21] S. Haddad, N. Pekergin, Using stochastic comparison for efficient model checking of uncertain markov chains, in: *Quantitative Evaluation of Systems, QEST'09*, IEEE Computer Society, 2009, pp. 177–186.
- [22] E.M. Hahn, H. Hermanns, B. Wachter, L. Zhang, Time-bounded model checking of infinite-state continuous-time Markov chains, *Fund. Inform.* 95 (2009) 129–155.
- [23] E.M. Hahn, H. Hermanns, B. Wachter, L. Zhang, PASS: abstraction refinement for infinite probabilistic models, in: J. Esparza, R. Majumdar (Eds.), *Tools and Algorithms for the Construction and Analysis of Systems, TACAS 2010*, in: LNCS, vol. 6015, Springer, 2010, pp. 353–357.
- [24] H. Hansson, B. Jonsson, A logic for reasoning about time and reliability, *Form. Asp. Comput.* 6 (1994) 512–535.
- [25] H. Hermanns, B. Wachter, L. Zhang, Probabilistic CEGAR, in: A. Gupta, S. Malik (Eds.), *Computer Aided Verification, CAV 2008*, in: LNCS, vol. 5123, Springer, 2008, pp. 162–175.
- [26] M. Huth, On finite-state approximants for probabilistic computation tree logic, *Theoret. Comput. Sci.* 246 (2005) 113–134.
- [27] B. Jonsson, K.G. Larsen, Specification and refinement of probabilistic processes, in: *Annual Symposium on Logic in Computer Science, LICS 1991*, IEEE Computer Society, 1991, pp. 266–277.
- [28] J.P. Katoen, D. Klink, M. Leucker, V. Wolf, Three-valued abstraction for continuous-time Markov chains, in: W. Damm, H. Hermanns (Eds.), *Computer Aided Verification, CAV 2007*, in: LNCS, vol. 4590, Springer, 2007, pp. 311–324.
- [29] J. Katoen, D. Klink, M. Leucker, V. Wolf, Three-valued abstraction for probabilistic systems, *Tech. Report AIB-2007-20*, RWTH Aachen, Germany, 2007.
- [30] M. Kattenbelt, M.Z. Kwiatkowska, G. Norman, D. Parker, Game-based probabilistic predicate abstraction in PRISM, in: *Proceedings of the Sixth Workshop on Quantitative Aspects of Programming Languages, QAPL 2008*, in: ENTCS, vol. 220(3), Elsevier, 2008, pp. 5–21.
- [31] R.B. Kearfott, Interval computations: introduction, uses, and resources, *Euromath Bull.* 2 (1996) 95–112.
- [32] D. Klink, Three-valued abstraction for stochastic systems, Ph.D. Thesis, RWTH Aachen University, Germany, 2010.
- [33] I. Kozine, L.V. Utkin, Interval-valued finite markov chains, *Reliab. Comput.* 8 (2002) 97–113.
- [34] M.Z. Kwiatkowska, G. Norman, D. Parker, PRISM: probabilistic symbolic model checker, in: T. Field, P.G. Harrison, J.T. Bradley, U. Harder (Eds.), *Computer Performance Evaluation, Modelling Techniques and Tools, TOOLS 2002*, in: LNCS, vol. 2324, Springer, 2002, pp. 200–204.
- [35] M.Z. Kwiatkowska, Model checking for probability and time: from theory to practice, in: *IEEE Symposium on Logic in Computer Science, LICS 2003*, IEEE Computer Society, 2003, pp. 351–360.
- [36] M.Z. Kwiatkowska, G. Norman, D. Parker, Game-based abstraction for Markov decision processes, in: *Quantitative Evaluation of Systems, QEST 2006*, IEEE Computer Society, 2006, pp. 157–166.
- [37] M.B. Mamoun, N. Pekergin, Model checking of infinite state space Markov chains by stochastic bounds, in: K. Al-Begain, A. Heindl, M. Telek (Eds.), *Analytical and Stochastic Modeling Techniques and Applications, ASMTA 2008*, in: LNCS, vol. 5055, Springer, 2008, pp. 264–278.
- [38] D. Monniaux, Abstract interpretation of programs as Markov decision processes, *Sci. Comput. Programming* 58 (2005) 179–205.
- [39] G. Păun, *Membrane Computing. An Introduction*, Springer-Verlag, 2002.
- [40] K. Sen, M. Viswanathan, G. Agha, Model-checking markov chains in the presence of uncertainties, in: H. Hermanns, J. Palsberg (Eds.), *Tools and Algorithms for the Construction and Analysis of Systems, TACAS 2006*, in: LNCS, vol. 3920, Springer, 2006, pp. 394–410.
- [41] D. Skulj, Finite discrete time Markov chains with interval probabilities, in: J. Lawry, E. Miranda, A. Bugarin, S. Li, M.A. Gil, P. Grzegorzewski, O. Hryniewicz (Eds.), *Soft Methods for Integrated Uncertainty Modelling, SMPS 2006*, in: *Advances in Soft Computing*, vol. 37, Springer, 2006, pp. 299–306.
- [42] D. Skulj, Discrete time markov chains with interval probabilities, *Internat. J. Approx. Reason.* 50 (2009) 1314–1329.
- [43] M. Villasana, A. Radunskaia, A delay differential equation model for tumor growth, *J. Math. Biol.* 47 (2003) 270–294.
- [44] B. Wachter, L. Zhang, H. Hermanns, Probabilistic model checking modulo theories, in: *Fourth International Conference on the Quantitative Evaluation of Systems, QEST 2007*, IEEE Computer Society, 2007, pp. 129–140.
- [45] K. Weichselberger, The theory of interval-probability as a unifying concept for uncertainty, in: G. De Cooman, F. Gagliardi Cozman, S. Moral, P. Walley (Eds.), *Imprecise Probabilities and their Applications, ISIPTA 1999*, pp. 387–396.